

热点追踪

三维人脸精细结构重建

王子都 朱翔昱 雷震

中国科学院自动化研究所

本文主要介绍中国科学院自动化研究所生物识别与安全技术研究中心在三维人脸精细结构重建领域的最新研究成果，包括发表在CVPR 2024的3DDFA-V3^[1]和ACM MM 2024的S2TD-Face^[2]，并围绕其技术细节和创新点展开讨论。

一、研究背景

三维人脸重建作为计算机视觉与图形学领域的一个重要的方向，在生物识别、影视娱乐、人机交互、医疗美容等具有广阔的应用前景。近年来，基于深度学习的三维人脸重建方法利用数据驱动和模型先验，能够从非受控条件下的人脸图片中建模三维人脸几何形状和纹理信息，典型方法如 3DDFA 系列^{[3][4]}、DECA^[5]、Deep3D^[6]、HRN^[7]等。然而，三维人脸的精细结构重建在许多应用中仍然面临巨大挑战，并具有不可替代的重要性。例如，在数字人驱动中，精细的人脸形状结构

建模直接影响虚拟人脸的真实感和自然度；在精神状态分析中，微表情和面部细节的准确捕捉对心理状态的推测具有重要价值。为进一步推动这一领域的发展，本文探讨了三维人脸精细结构重建中的两个具体研究工作，旨在抛砖引玉，助力相关领域的实际应用。

二、部件分割引导的人脸细节捕捉

2.1 背景介绍

从二维图像中重建三维人脸是计算机视觉研究的一项关键任务，研究人员通常依赖三维可变形模型 (3DMM) 进行人脸重建，以定位面部特征和捕捉表情。如图 1 所示，现有的方法往往难以准确重建出如闭眼、歪嘴、皱眉等极端表情，并且三维误差有时难以精确描述二维区域的对齐情况。为了增强 3DMM 对极端表情的捕捉和重建能力，3DDFA-V3 从训练策略和数据策略两个角度进行研究，以人脸分割为研究切入点，使用人脸部件分割的几何信息作为监督信号，设计损失函数，显著加强了对形状的约束，同时，3DDFA-V3 设计了可靠的表情生成方法，能够大批量、可控地生成难以获取的极端表情人脸图像。

2.2 研究内容

研究内容 1: 训练策略的研究

人脸部件分割能够以像素级的精度为每个面部特征提供准确的定位。相比常用的关键点信息，部件分割提供了覆盖整个图像的更密集的标签；相比纹理信息，部件分割不易受到光照等因素的干扰。如图 2 所示，3DDFA-V3 的总体思路是利用分割信息来直接指导三维形变，进一步增强对人脸形状的约束。

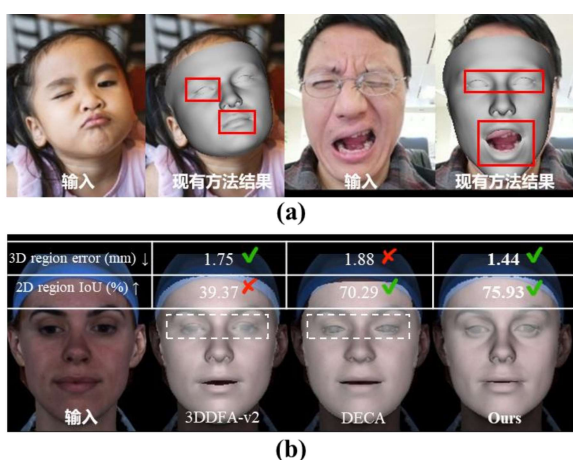


图 1 (a) 现有方法难以重建闭眼、歪嘴等极端表情。
(b) 三维误差有时难以精确描述二维区域的对齐情况。

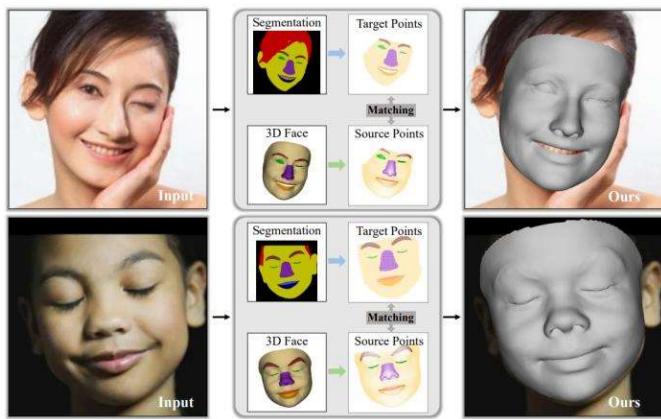


图 2 3DDFA-V3 的总体思路

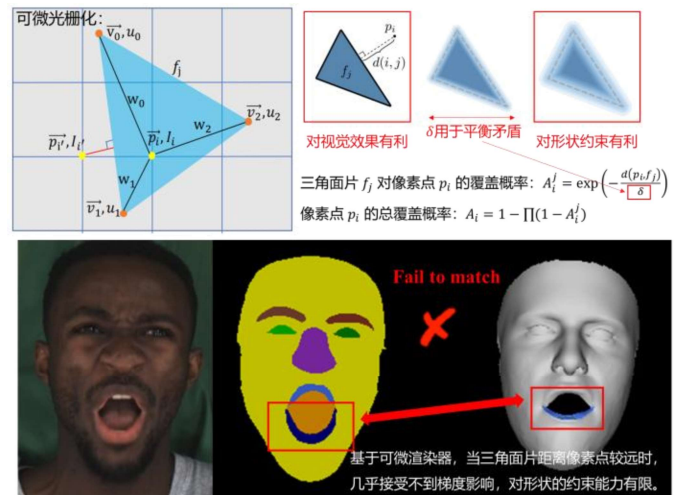


图 3 可微渲染器对形状约束能力不强的原因分析

现有利用分割信息来拟合三维人脸的尝试，依赖于可微渲染器生成预测的面部区域（如眼睛、眉毛、嘴唇等）轮廓，并使用类似 IoU 的损失函数优化渲染轮廓与分割之间的差异。在这一过程中，可微渲染器是梯度传播的媒介，其关键步骤被称作可微光栅化（Differentiable Rasterization），把图片的 Alpha channel 变成了一个受像素到 Mesh 中三角面片的距离影响的概率分布图。3DDFA-V3 简单讨论了“距离影响概率”这一可微渲染的基本设定可能存在问题：一方面，为了让产生的梯度有效地影响三维形变，被渲染的像素点应该受到全局的三角面片的纹理等属性的影响；另一方面，从视觉渲染效果角度出发，为了让渲染像素点的纹理足够清晰，只能让其尽可能只受到离它最近的三角面片的影响；两者存在矛盾，导致三角面片距离像素点较远时，几乎接受不到梯度影响，如图 3 所示。因此，3DDFA-V3 认为可微渲染器对形状约束的能力不强，希望利用分割信息设计更直接有效的损失函数引导三维人脸形变。

3DDFA-V3 的关键思想是将目标和预测的部件分割转化为语义点集，通过优化点集的分布来确保重建区域和目标具有相同的几何形态。具体来讲，3DDFA-V3 提出了部件重投影距离损失（Part Re-projection Distance Loss, PRDL）。PRDL 按照区域 $P = \{\text{left-eye, right-eye, left-eyebrow, right-eyebrow, up-lip, down-lip, nose, skin}\}$ 对人脸进行分块，针对二维部件分割的每个部分 $p \in P$ ，PRDL 首先在分割区域内采样点，得到目标点集 $\{C_p \mid p \in P\}$ 。然后，PRDL 将三维人脸重建结果重新投影到图像平面上，并根据人脸模型的 masks 获得与目标区域 $\{C_p \mid p \in P\}$ 语义一致的预测点集 $\{V_{2d}^p(\alpha) \mid p \in P\}$ ， α 是人脸模型的系数。接着 PRDL 对图像平面的网格点进行采样，得到锚点集合 A ，并计算任意一个锚点 $a_i \in A$ 到点集的各种统计距离（如最近距离 f_{min} 、最远距离 f_{max} 、平均距离 f_{ave} 等）来建立几何描述子。最后，PRDL 通过优化相同语义的预测点集的几何

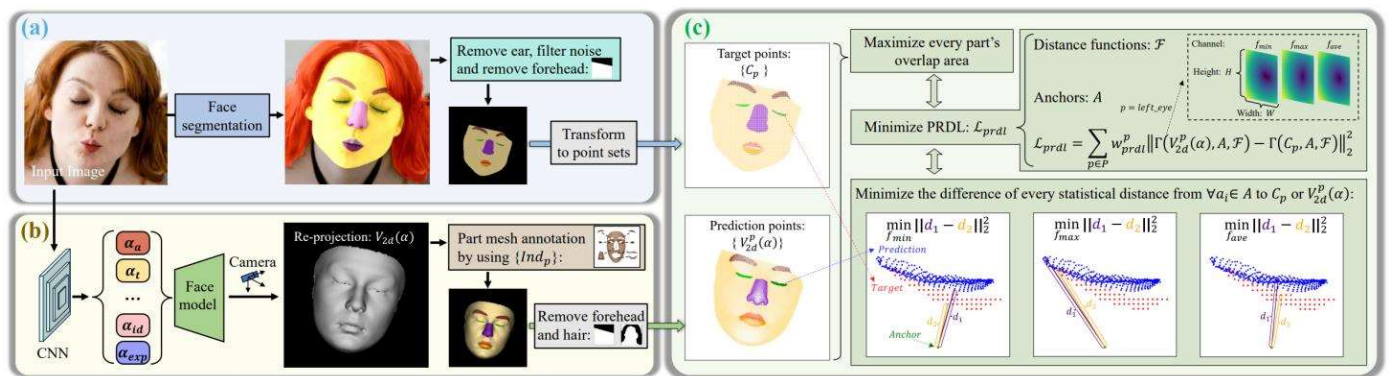


图 4 部件重投影距离损失（PRDL）概述，其核心思想是最小化每个锚点到目标和预测点集的统计距离的差异

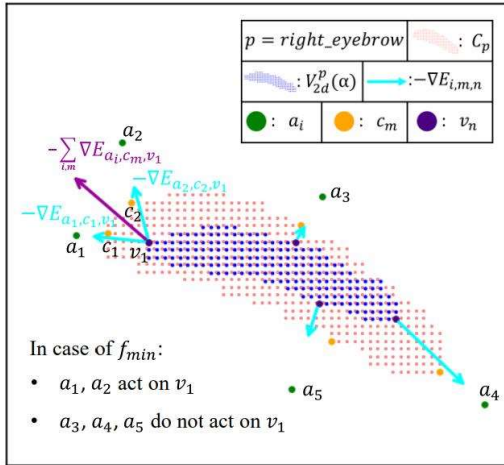


图 5 PRDL 的梯度分析图, PRDL 的锚点 a_i 对预测点有选择以及“推”和“拉”的作用, 最终的作用在预测点上的梯度是相关锚点作用的叠加, 具有鲁棒性。

描述子和目标点集的几何描述子的差异, 确保重建区域和目标具有相同的几何分布, 从而提高目标和预测点集覆盖区域之间的重叠度, 整个过程如图 4 所示。概括来讲, 对于 $\forall p \in P$ 和 $\forall a_i \in A$, PRDL 的优化目标为:

$$\text{For } \forall p \in P, \forall a_i \in A: \begin{cases} \min \|f_{min}(V_{2d}^p(\alpha), a_i) - f_{min}(C_p, a_i)\|_2 \\ \min \|f_{max}(V_{2d}^p(\alpha), a_i) - f_{max}(C_p, a_i)\|_2 \\ \min \|f_{ave}(V_{2d}^p(\alpha), a_i) - f_{ave}(C_p, a_i)\|_2 \end{cases}$$

在使用过程中, PRDL 可以使用最远点采样 (Farthest Point Sampling) 等技术可以减少 C_p 、 $V_{2d}^p(\alpha)$ 和 A 中的点数量, 从而降低计算成本。通过理论推导可知, 在梯度下降的条件下, PRDL 的锚点 a_i 对预测点有选择以及“推”和“拉”的作用, 最终能使得锚点 a_i 到预测点集和目标点集的统计距离尽可能相同, 从而指导三维人脸的形变, 如图 5 所示, 当预测点被目标点包围

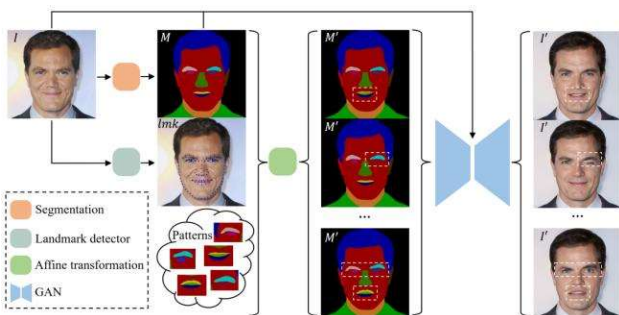


图 6 可控、大批量地合成闭眼、歪嘴、皱眉等表情

时, 锚点可以对预测点产生向外扩展的梯度下降方向。

研究内容 2: 数据策略的研究

虚拟合成数据常用于训练三维人脸重建模型。现有的合成人脸数据要么侧重于背景、光照和身份的多样化, 要么集中在姿态变化上, 虽然在重建自然面部表情方面提供了有效的帮助, 但在重建极端表情方面难以提供支持。为了克服这些局限并促进相关研究, 3DDFA-V3 采用了一种基于 GAN 的方法来大批量可控地合成难以搜集的人脸极端表情数据, 包括闭眼、歪嘴和皱眉等表情。

如图 6 所示, 我们首先使用人脸分割方法和关键点检测器分别获取原始图像 I 的分割图 M 和关键点 lmk , 并预设一些人脸表情的局部变化模板 $Patterns$ 。利用关键点 lmk 的位置信息, 对 $Patterns$ 进行合适的仿射变换, 将其应用到原始分割图 M 上, 得到 M' 。随后将 M' 输入到一个条件 GAN 网络中, 生成新的面部表情图像 I' 。目前 3DDFA-V3 已经生成了超过 50 万张的闭眼、歪嘴、皱眉等表情的数据, 并进行了开源, 数据示例如图 7 所示。

2.3 实验分析与结果

详细的实验设置、定量或定性实验可参考 3DDFA-V3 论文, 下面仅对关键实验进行展示。

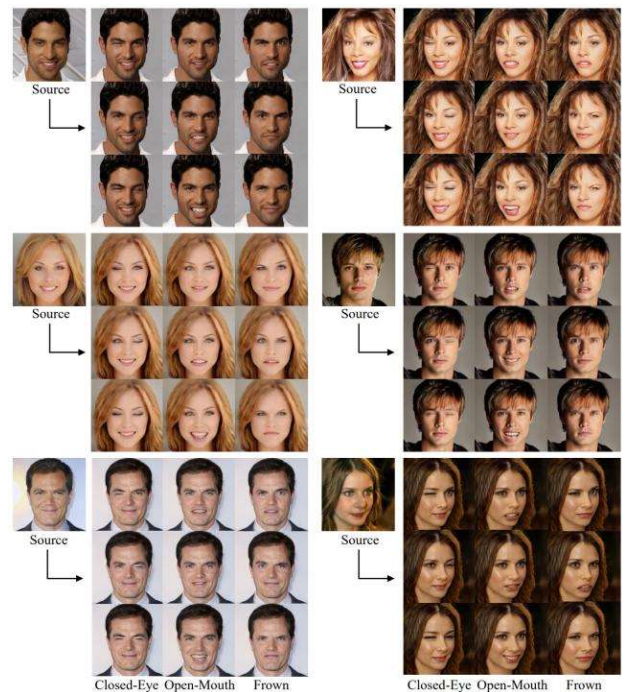


图 7 3DDFA-V3 提供的虚拟合成表情数据示例

Methods	Frontal-view (mm) ↓					Side-view (mm) ↓				
	Nose	Mouth	Forehead	Cheek	avg.	Nose	Mouth	Forehead	Cheek	avg.
	avg.± std.	avg.± std.	avg.± std.	avg.± std.		avg.± std.	avg.± std.	avg.± std.	avg.± std.	
PRNet ^[8]	1.923±0.518	1.838±0.637	2.429±0.588	1.863±0.698	2.013	1.868±0.510	1.856±0.607	2.445±0.570	1.960±0.731	2.032
MGCNet ^[9]	1.771±0.380	1.417±0.409	2.268±0.503	1.639±0.650	1.774	1.827±0.383	1.409±0.418	2.248±0.508	1.665±0.644	1.787
Deep3D ^[6]	1.719±0.354	1.368±0.439	2.015±0.449	1.528±0.501	1.657	1.749±0.343	1.411±0.395	2.074±0.486	1.528±0.517	1.691
3DDFA-v2 ^[4]	1.903±0.517	1.597±0.478	2.447±0.647	1.757±0.642	1.926	1.883±0.499	1.642±0.501	2.465±0.622	1.781±0.636	1.943
HRN ^[7]	1.722±0.330	1.357±0.523	1.995±0.476	1.072±0.333	1.537	1.642±0.310	1.285±0.528	1.906±0.479	1.038±0.322	1.468
DECA ^[5]	1.694±0.355	2.516±0.839	2.394±0.576	1.479±0.535	2.010	1.903±1.050	2.472±1.079	2.423±0.720	1.630±1.135	2.107
Ours	1.586±0.306	1.238±0.373	1.810±0.394	1.111±0.327	1.436	1.623±0.313	1.205±0.366	1.864±0.424	1.076±0.315	1.442

表 1 3DDFA-V3 在 REALY benchmark 上取得了 SOTA 的水平

定量对比实验: 如表 1 所示, 3DDFA-V3 在 REALY benchmark 上取得了 SOTA 的水平。

定性对比实验: 如图 8 所示, 与现有的 SOTA 方法相比, 3DDFA-V3 的预测结果可以非常准确地重建不对称和奇怪的面部表情。

三、基于素描草图的带纹理三维人脸重建

3.1 背景介绍

从人脸素描草图中重建带有纹理的精细 3D 人脸在刑侦与失踪人员调查、动漫娱乐、艺术设计等多个场景中具有广泛的应用潜力, 是一个极具前景但尚未充分发



图 8 与现有的 SOTA 方法相比, 3DDFA-V3 的预测结果可以非常准确地重建不对称和奇怪的面部表情

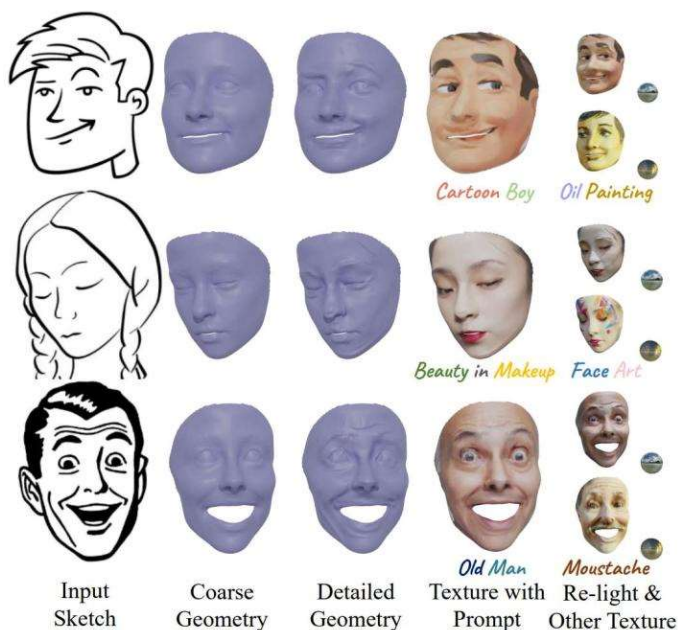


图 9 S2TD-Face 能够从不同风格的人脸草图中重建出高保真且拓扑一致的 3D 精细人脸

如图 9 所示，我们提出了一种从素描草图中重建具有可控纹理的 3D 人脸的新方法，称为 S2TD-Face (Sketch to controllable Textured and Detailed Three-Dimensional Face)。S2TD-Face 引入了一个两阶段形状重建框架，能够直接从输入草图中重建精细的带纹理的三维人脸形状。为了将素描的细节笔触反馈到重建的 3D 形状上，S2TD-Face 提出了一种新的草图到几何形状的损失函数，以确保重建结果精确匹配输入特征，如草图勾勒出的酒窝和皱纹等。S2TD-Face 的训练不依赖难以获取的 3D 人脸扫描数据或手绘素描草图。此外，S2TD-Face 还引入了一个纹理控制模块，通过文本提示从纹理库中选择合适的纹理并将其无缝整合到几何结构中，从而得到具有可控纹理的 3D 细节人脸，此外，S2TD-Face 还支持基于 ControlNet 的 3D 人脸纹理控制方法。

3.2 研究内容

图 10 是 S2TD-Face 的总体流程概括，S2TD-Face 主要包括几何形状重建模块和纹理控制模块。

研究内容 1: 训练策略的研究

基于现有大量的二维真实人脸图片，S2TD-Face 首先集成了各种素描草图生成方法，从二维人脸图片中得

展的研究领域。现有研究主要面临着两方面的不足：一方面，现有的方法只能处理姿态受限且有真实阴影的人脸素描草图，且难以将素描的细节笔触反馈到重建的 3D 形状上；另一方面，纹理在面部外观的表现中起着关键作用，但素描草图缺乏这一信息，因此在重建过程中需要额外的纹理控制。

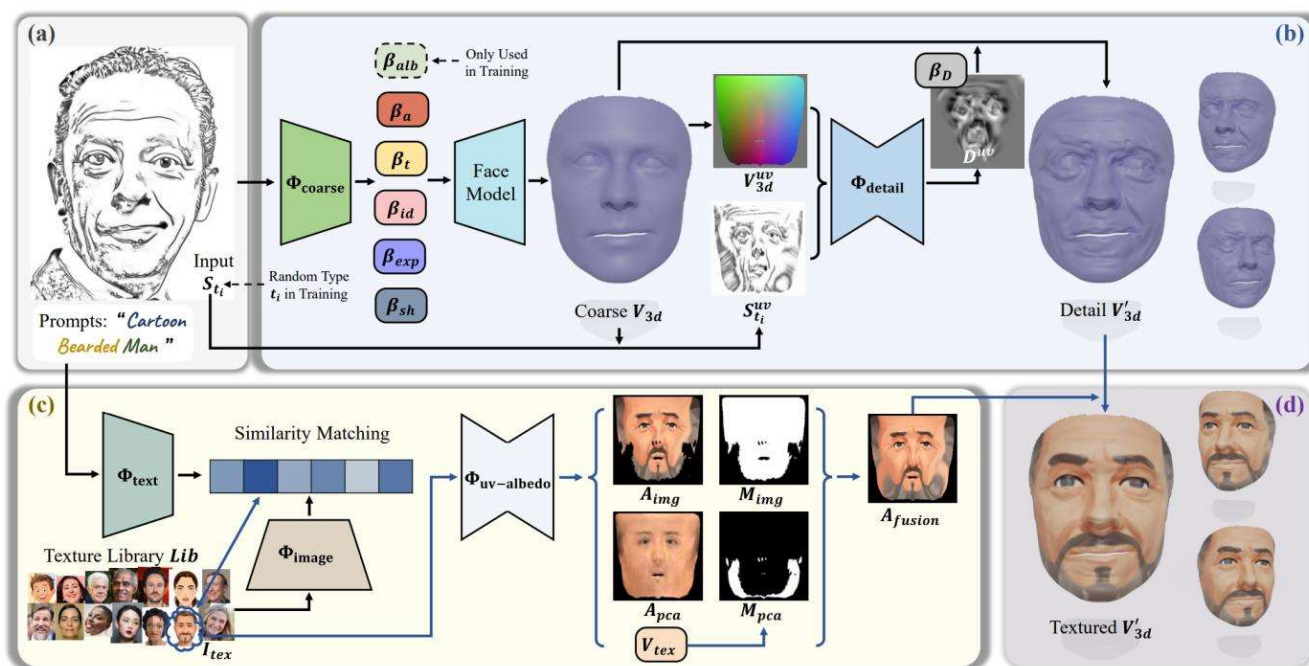


图 10 基于素描输入的纹理可控的三维精细人脸重建方法的流程概括

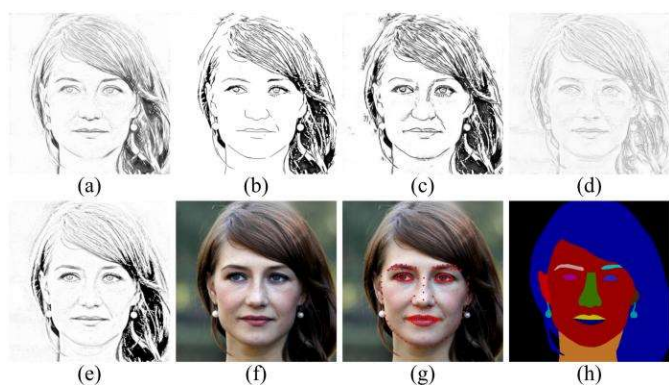


图 11 S2TD-Face 的数据示例

到了各类风格的素描人脸。由于每张素描人脸数据都有真实的二维人脸图片数据与之对应, S2TD-Face 在训练的时候能够结合成熟的三维人脸重建技术, 利用已有的关键点信息、五官分割信息和纹理信息对重建出的 3D 形状进行约束。这种训练策略使得 S2TD-Face 不依赖于难以收集的 3D 人脸扫描数据和手绘草图。图 11 是 S2TD-Face 的数据示例, (a)-(e)为从原始图像(f)中生成的不同风格的草图, (g)表示关键点, (h)表示分割信息。S2TD-Face 重建框架的输入包括素描草图 (a)-(e), (f)-(h)用作监督信号。

研究内容 2: 重建与素描细节笔触一致的精细三维人脸

基于素描草图的特点, S2TD-Face 结合可微渲染技术, 设计了有效的损失函数, 其能够捕捉素描草图刻画的人脸形状信息, 并将其准确地重建到三维结构上。S2TD-Face 使用法线偏移对人脸 mesh 进行精细化建模, 并构建素描到三维信息的损失函数 $\mathcal{L}_{\text{sketch}}$:

$$\mathcal{L}_{\text{sketch}} = \lambda_1 \underbrace{\sum_{n \in \{a,b,c,d\}} \|M^n - M\|_2}_{\text{sketch-photometric}} + \lambda_2 \underbrace{\sum_{n \in \{a,b,c,d\}} \left(1 - \frac{\langle \Phi_{\text{per}}(M^n), \Phi_{\text{per}}(M) \rangle}{\|\Phi_{\text{per}}(M^n)\|_2 \cdot \|\Phi_{\text{per}}(M)\|_2} \right)}_{\text{sketch-perception}}$$

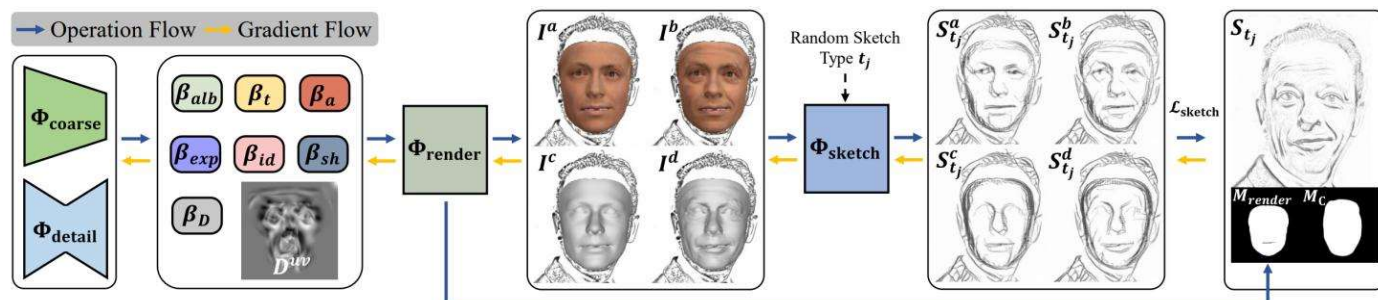


图 12 素描到三维信息的损失函数 $\mathcal{L}_{\text{sketch}}$ 的过程概述

其中, M^n 和 M 分别表示预测的素描图与对应真值素描图的经过面部区域掩码过滤的结果, $n \in \{a, b, c, d\}$ 表示预测素描图的四种形式, 即 a 是由纹理和粗糙形状渲染得到的预测素描图; b 是由纹理和精细形状渲染得到的预测素描图; c 是由灰色纹理和粗糙形状渲染得到的预测素描图; d 是由灰色纹理和精细形状渲染得到的预测素描图。 $\mathcal{L}_{\text{sketch}}$ 包括两个部分, 第一部分计算渲染预测的素描与对应真值的图片度量损失 (sketch-photometric), 第二部分计算渲染预测的素描与对应真值的感知度量损失 (sketch-perception), $\mathcal{L}_{\text{sketch}}$ 的可视化过程如图 12 所示, 更详细的计算过程也可参考 S2TD-Face 原文。

研究内容 3: 精细三维人脸纹理控制模块

对于精细三维人脸纹理控制模块, S2TD-Face 首先搜集一定数量的各种外观风格的人脸图像, 作为预设的人脸纹理模板库 Library。S2TD-Face 将使用者提供的待重建三维人脸的纹理的文本描述 Text, 作为纹理控制模块的输入。利用 CLIP 对预设的人脸纹理模板库 Library 中的图片进行匹配, 选取最相似的人脸纹理图片或从最相似的前 k 张图片中选取任意一张图片, 保证方法的灵活性, 利用三维人脸重建技术估计人脸图片的 UV 纹理展开图, 并同时估计出 PCA 纹理对不可见区域进行补全。此外, 最新开源的 S2TD-Face 还进一步支持了基于 ControlNet 的 3D 人脸纹理控制方法。

3.3 实验分析与结果

详细的实验设置、定量或定性实验可参考 S2TD-Face 论文, 下面仅对关键实验进行展示。

可视化结果: 如图 13 所示, S2TD-Face 能够从不

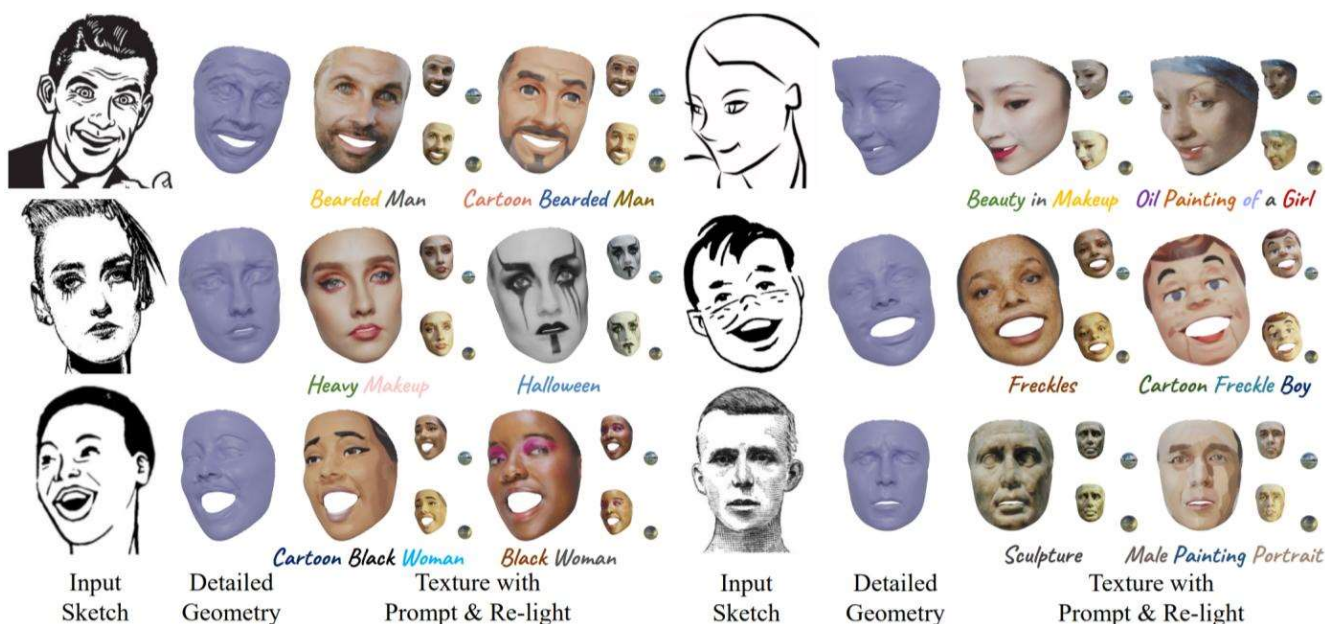


图 13 S2TD-Face 的可视化结果

对比实验: 如图 14 所示, 与现有方法相比, S2TD-Face 的重建结果展现出了与输入人脸素描草图细节和身份高度一致的最佳效果。

3D 形状细节的捕捉方式: 如图 15 所示, 得益于有效的几何形状重建模块和纹理控制模块的独立设计, S2TD-Face 对人脸素描草图刻画细节的捕捉方式是基于 3D 精细形状的, 不受纹理变化的影响。

四、总结展望

三维人脸精细结构重建作为三维人脸领域的重要方向, 对于数字人驱动、精神状态分析等实际应用具有深远意义。尽管现有方法在精细化建模方面取得了显著进展, 但仍存在局部细节捕捉不足、动态表情重建受限等挑战, 此外, 与可微渲染流程结合设计更有力的形状约束方式也是值得探讨的问题。本文以三维人脸精细结构重建为主题, 讨论了基于部件分割引导的人脸细节捕捉 (3DDFA-V3) 和基于素描草图的带纹理三维人脸重建 (S2TD-Face) 这两项科研工作, 希望为该领域的进一步发展提供借鉴和启发。本文介绍的 3DDFA-V3 和 S2TD-Face 的代码请参考 <https://github.com/wang-zidu/3DDFA-V3> 和 <https://github.com/wang-zidu/S2TD-Face>。

责任编辑 张杰

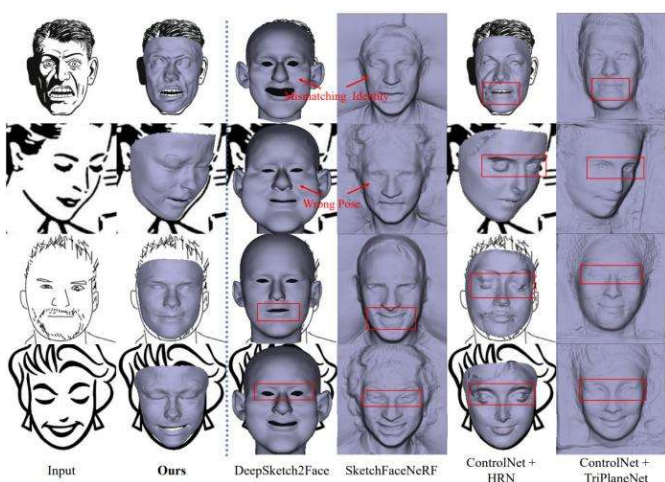


图 14 S2TD-Face 与现有方法对比

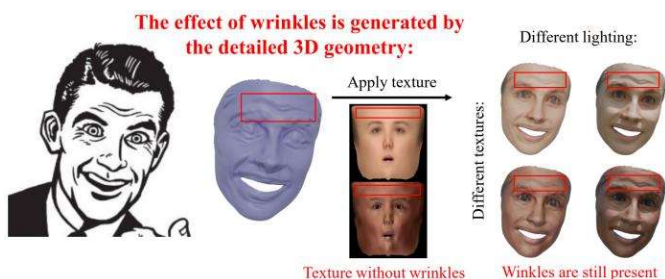


图 15 S2TD-Face 的局部细节效果 (如皱纹、酒窝等) 是由形状表示的, 不依赖于特定的纹理

同风格的人脸草图中重建出高保真且拓扑一致的 3D 精细人脸。它还支持基于文本提示的 3D 人脸纹理控制, 能够生成卡通、雕塑风格或真实人脸风格的纹理。

参考文献

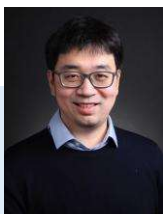
- [1] Wang, Zidu, et al. 3D Face Reconstruction with the Geometric Guidance of Facial Part Segmentation. In CVPR 2024.
- [2] Wang, Zidu, et al. S2TD-Face: Reconstruct a Detailed 3D Face with Controllable Texture from a Single Sketch. In ACM MM 2024.
- [3] Zhu, Xiangyu, et al. Face alignment in full pose range: A 3d total solution. In IEEE T-PAMI 2017.
- [4] Guo, Jianzhu, et al. Towards fast, accurate and stable 3d dense face alignment. In ECCV 2020.
- [5] Feng, Yao, et al. Learning an animatable detailed 3D face model from in-the-wild images. In ToG 2021.
- [6] Deng, Yu, et al. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In CVPR workshops 2019.
- [7] Lei, Biwen, et al. A hierarchical representation network for accurate and detailed face reconstruction from in-the-wild images. In CVPR 2023.
- [8] Feng, Yao, et al. Joint 3d face reconstruction and dense alignment with position map regression network. In ECCV 2018.
- [9] Shang, Jiayang, et al. Self-supervised monocular 3d face reconstruction by occlusion-aware multi-view geometry.



王子都

王子都，中国科学院自动化研究所 2022 级硕士研究生，导师为朱翔昱副研究员，主要研究方向为三维人脸重建。

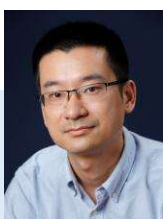
Email: wangzidu2022@ia.ac.cn



朱翔昱

朱翔昱，中国科学院自动化研究所副研究员，从事生物特征识别、数字人、三维重建等方向的理论与应用。国际模式识别协会（IAPR）生物特征青年学者奖（YBIA）获得者（两年一次，每次从全球范围内评选 40 岁以下学者一名）。任 CCF:A 类期刊 T-IFS Associate Editor。共发表论文 80 余篇，发表文章的 Google Scholar 总引用次数为 9600 余次。获得三次国际竞赛冠军以及四项最佳论文及提名奖。授权国家发明专利 12 项。入选 IEEE Senior Member，百度学术全球华人 AI 青年学者榜单（全球 25 人），受到腾讯犀牛鸟基金支持。获 2021 中国电子学会科技进步二等奖、中国图象图形学学会优秀博士论文提名奖。提出的人脸三维建模方法在被 PyTorch 官方 Twitter 报道，开源代码在 GitHub 上收获 6000 余星。

Email: xiangyu.zhu@nlpr.ia.ac.cn



雷震

雷震，IEEE/IAPR Fellow，中国科学院自动化研究所研究员，中国科学院大学岗位教授，中国科学院香港创新院人工智能与机器人创新中心教授，博士生导师。致力于人工智能基础理论，图像视频分析与理解，生物特征识别方面的研究。Google Scholar 文章引用次数 3.3 万余次，H-index: 85，2020-2023 爱思唯尔中国高被引学者，授权发明专利 30 余项，撰写发布国家标准 2 项，国家公共安全行业标准 7 项，十余次获得国际学术会议最佳（学生）论文奖或国际会议竞赛第一名。

Email: zhen.lei@ia.ac.cn